UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY · DAVIS · IRVINE · LOS ANGELES · RIVERSIDE · SAN DIEGO · SAN FRANCISCO

SANTA BARBARA · SANTA CRUZ

Space Sciences Laboratory

BERKELEY, CALIFORNIA 94720

Tel: (510) 642-2969
e-mail: westphal@ssl.berkeley.edu
January 25, 2016

Response of the CAPTEM Informatics Sub-committee to RFI NNH15ZDA012L: Preparation for the development of a community-based roadmap for NASAs Planetary Data Services

Here we address question 4: What role should the PDS play, relative to other archiving alternatives (including scientific journals), in providing the public access to the data that is the product of NASAs funded research and the basis of published scientific studies?

In recognition of the long-term need to make the results of publicly-funded research available to the public, and in response to the directive from OSTP, the Informatics Subcommittee of the Curation and Planning Team for Extraterrestrial Materials is evaluating different methods for making data publicly available, including the Planetary Data System. In addition, they are evaluating requirements on proprietary data. These issues are on the agenda for discussion at the Spring 2016 CAPTEM meeting.

NASA currently curates eight distinct collections of ET materials: Apollo lunar samples, meteorites collected in Antarctica, Interplanetary Dust Particles (IDPs) collected in the stratosphere, collectors with embedded solar wind from the Genesis mission, Stardust returned samples from comet Wild 2, the Interstellar Dust Collector also from the Stardust mission, and samples returned by the Japanese Hayabusa mission, and impacts on Returned Space Hardware, including the LDEF mission.

Laboratory analyses of extraterrestrial (ET) materials address fundamental questions in planetary science. This is a very dynamic field: laboratory instruments are continuously being modified and improved. They are often home-grown by PIs with limited resources. The truly astounding improvements in laboratory instrumentation since Apollo, wisely supported by NASA starting in the early 1960s, are something of which NASA should be incredibly proud. These improvements have benefitted scientific research generally, not just in the study of ET materials.

Although the PDS now accepts sample data, other geochemical databases exist that may be better suited than the PDS for this purpose. The PDS was originally designed for archiving of data and data products from telemetry from planetary missions. These data are well-defined and because of bandwidth limitations are relatively small in size. In contrast, data collected in laboratories are not so well-defined. We believe that a number of significant issues must be addressed before laboratory analyses of returned ET samples should be considered for incorporation into the PDS.

- There are of order 30 types of instruments commonly used in ET sample analyses. Among each type, there are often several different specific instruments. Some are commercial, and some are unique and home-grown by PIs. The result is a bewildering plethora of instruments in the community, and a hodge-podge of data formats. In almost no case has an effort been made to agree on a community-wide data format within instrument types, largely because there is little need for such a consensus. Such an effort would be possible, but the benefits of such an effort would likely be minimal as compared with the cost.
- Similarly, generally there is no community consensus on formats for data products.
- Without a community consensus, archived data and data products would have to be extensively documented to be of any value to current or future researchers. This documentation would have be written, evaluated, and maintained. In a rapidly changing field, maintenance would be a major challenge. There is currently no support for such activities.
- There is no community consensus on the types of well-documented data or data products that would be of use to current or future researchers. While it is likely possible technically (though very challenging) to archive all of the data collected by every instrument in the community, there is little value in archiving everything. In practice, few analyses proceed like data collection by an instrument on a spacecraft: there are aborted or failed analyses, equipment failures, risky analyses that are not successful, and so on. Data analyses are similarly often a work in progress, and there is usually not a clear criterion for deciding which among the various levels of data analysis should be archived. Even if such a consensus could be developed, it would be highly specific to the type of analysis – a generic requirement is probably impossible to write. All these data and data products could in principle be collected, but to be of any value they would have to be documented, which would be time-consuming and would discourage experimentation and risk-taking.

We recognize the need to make data collected on the public dime publicly accessible. We point out that this has been happening for centuries, but only selectively, through publication of results in conference proceedings and peer-reviewed journals. The emerging requirement by journals to include data, data products and analysis procedures or code, and to include these in the review process, is a step forward and addresses many of the issues enumerated above.

We conclude that archiving of data and data products from laboratory analyses at the PI level on PDS is at best premature. Significant resources would be required to address the issues we have enumerated here. While it is possible in principle to address these issues through additional support, we are deeply concerned that an unfunded or inadequately funded mandate to adhere to new archiving requirements will dramatically slow research and throw a monkey-wrench into an extremely productive program that has delivered enormous benefits to society generally. We recommend that an assessment of existing geochemical databases, which may be more suitable for this purpose, be carried out. We also recommend that for the time being, if PDS is used for archiving data, it is restricted to supporting data,

data products, and code in support of peer-reviewed publications.

Sincerely,

CAPTEM Informatics Subcommittee
Andrew Westphal, U. C. Berkeley, Chair
Denton Ebel, American Museum of Natural History
Timothy McCoy, Smithsonian Institution